Supplementary Material

Abbas Kazemipour, Ji Liu, Krystyna Solarana, Daniel A. Nagode, Patrick O. Kanold, Min Wu and Behtash Babadi

I. Experimental Procedures

Surgery: 2 hours before surgery, 0.1 cc dexamethasone (2 mg/ml, VetOne) was injected subcutaneously to reduce brain swelling during craniotomy. Anesthesia is induced with 4% isoflurane (Fluriso, VetOne) with a calibrated vaporizer (Matrx VIP 3000). During surgery, isoflurane level was reduced to and maintained at a level of 1.5%–2%. Body temperature of the animal is maintained at 36.0 degrees Celsius during surgery. Hair on top of head of the animal was removed using Hair Remover Face Cream (Nair), after which Betadine (Purdue Products) and 70% ethanol was applied sequentially 3 times to the surface of the skin before removing the skin. Soft tissues and muscles were removed to expose the skull. Then a custom designed 3D printed stainless headplate was mounted over left auditory cortex and secured with C&B-Metabond (Parkell). A craniotomy with a diameter of around 3.5 mm was then performed over left auditory cortex. A three layered cover slip was used as cranial window, which is made by gluing (NOA71, Norland Products) 2 pieces of 3 mm coverslips (64-0720 (CS-3R), Warner Instruments) with a 5 mm coverslip (64–0700 (CS-5R), Warner Instruments). Cranial window was quickly dabbed in kwik-sil (World Precision Instruments) before mounted 3 mm coverslips facing down onto the brain. After kwik-sil cured, Metabond was applied to secure the position of the cranial window. Synthetic Black Iron Oxide (Alpha Chemicals) was then applied to the hardened Metabond surface. 0.05 cc Cefazolin (1 gram/vial, West Ward Pharmaceuticals) was injected subcutaneously when entire procedure was finished. After the surgery the animal was kept warm under heat light for 30 minutes for recovery before returning to home cage. Medicated water (Sulfamethoxazole and Trimethoprim Oral Suspension, USP 200 mg/40 mg per 5 ml, Aurobindo Pharms USA; 6 ml solution diluted in 100 ml water) substitute normal drinking water for 7 days before any imaging was performed.

Awake two-photon imaging: Spontaneous activity data of population of layer 2/3 auditory cortex (A1) neurons is collected from adult (3-month old) Thy1-GCaMP6s female mouse implanted with chronic window following the above procedure, using two-photon imaging. Acquisition is performed using a two-photon microscope (Thorlabs Bscope 2) equipped with a Vision 2 Ti:Sapphire laser (Coherent), equipped with a GaAsP photo detector module (Hamamatsu) and resonant scanners enabling faster high-resoluation scanning at 30–60 Hz per frame. The excitation wavelength was 920 nm. Regions ($\sim 300\ \mu\text{m}^2$) within A1 were scanned at 30 Hz through a 20x, 0.95 NA water-immersion objective (Olympus). During imaging the animal was head-fixed and awake. The microscope was rotated 45 degrees and placed over the left A1 where window was placed. An average image of field of view was generated by choosing a time window where minimum movement of the brain was observed and used as reference image for motion correction using TurboReg plugin in ImageJ. GCaMP6s positive cells are selected manually by placing a ring like ROI over each identified cell. Neuropil masks were generated by placing a 20 μm radius circular region over each cell yet excluding all cell soma regions. Traces of soma and neuropil were generated by averaging image intensity within respective masks at each time point. A ratio of 0.7 was used to correct for neuropil contamination.

Cell-attached patch clamp recordings and two-photon imaging: Recordings were performed in vitro in voltage clamp to simultaneously measure spiking activity and $\Delta F/F$. Thalamocortical slices containing A1 were prepared as previously described [1]. The extracellular recording solution consisted of artificial cerebral spinal fluid (ACSF) containing: 130 NaCl, 3 KCl, 1.25 KH2PO4, 20 NaHCO3, 10 glucose, 1.3 MgSO4, 2.5 CaCl2 (pH 7.35-7.4, in 95% O2 5% CO2). Action potentials were recorded extracellularly in loose-seal cell-attached configuration (seal resistance typically 20–30 MΩ) in voltage clamp mode. Borosilicate glass patch pipettes were filled with normal ACSF diluted 10%, and had a tip resistance of $\sim$ 3-5 MΩ in the bath. Data were acquired with a Multiclamp 700B patch clamp amplifier (Molecular Devices), low-pass filtered at 3-6 kHz, and digitized at 10 kHz using the MATLAB-based software. Action potentials were stimulated with a bipolar electrode placed in L1 or L23 to stimulate the apical dendrites of pyramidal cells (pulse duration 1-5 ms). Data were analyzed offline using MATLAB. Imaging was largely performed using a two-photon microscope (Ultima, Prairie Technologies) and a MaiTai DeepSee laser (SpectraPhysics), equipped with a GaAsP photo detector module (Hamamatsu) and resonant scanners enabling faster high-resoluation scanning at 30-60 Hz per frame. Excitation was set at 900 nm. Regions were scanned at 30 Hz through a 40x water-immersion objective (Olympus). Cells were manually selected as ring-like regions of interest (ROIs) that cover soma but exclude cell nuclei, and pixel intensity within each ROI was averaged to generate fluorescence over time and changes in fluorescence ($\Delta F/F$) were then calculated.

II. The Expectation Maximization Algorithm

In this section we give a short overview of the EM algorithm and its connection to iteratively re-weighted least squares (IRLS) algorithms. More details can be found in [2] and

A. Kazemipour, M. Wu, and B. Babadi are with the Department of Electrical & Computer Engineering, University of Maryland (UMD) , College Park, MD USA (e-mails: {kaazemi, minwu, behtash}@umd.edu). J. Liu, K. Solarana, Daniel A. Nagode, and P. O. Kanold are with the Department of Biology, UMD, College Park, MD USA (e-mails: liuji@umd.edu, korzecho@umd.edu, daniel.nagode@gmail.com, pkanold@umd.edu).

Corresponding authors: A. Kazemipour (kaazemi@umd.edu) and Behtash Babadi (behtash@umd.edu).

the references therein. Given the observations $\mathbf{y}$, the goal of the EM algorithm is to find the ML estimates of a set of parameters $\boldsymbol{\Theta}$ by maximizing the likelihood $\mathfrak{L}(\boldsymbol{\Theta}) := p(\mathbf{y}|\boldsymbol{\Theta})$. Such maximization problems are typically intractable, but often become significantly simpler by introducing a latent variable $\mathbf{u}$. The EM algorithm connects solving the ML problem to maximizing $\widetilde{\mathfrak{L}}(\boldsymbol{\Theta}) := p(\mathbf{y},\mathbf{u}|\boldsymbol{\Theta})$, if one knew $\mathbf{u}$.

Consider the state-space model:

$$\begin{aligned} \mathbf{x}_t &= \boldsymbol{\Theta}\mathbf{x}_{t-1} + \frac{\boldsymbol{\omega}_t}{\sqrt{\mathbf{u}_t}}, \\ \mathbf{y}_t &= \mathbf{A}_t\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}), \end{aligned} \tag{1}$$

where $\boldsymbol{\omega} \sim \mathcal{N}(0,\mathbf{I})$, $\mathbf{u}_t$ is a positive i.i.d. random vector, and the square root operator and division of the two vectors are understood as element-wise operations. Let $\delta^2_{t,j} := (\mathbf{x}_t - \boldsymbol{\Theta}\mathbf{x}_{t-1})^2_j$ for $j = 1,2,\cdots,p$. For an appropriate choice of the density of $(\mathbf{u}_t)_j$ denoted by $p_U(\cdot)$, we have [2]:

$$p(\tfrac{\boldsymbol{\omega}_t}{\sqrt{\mathbf{u}_t}}|\boldsymbol{\Theta}) = p(\mathbf{x}_t|\mathbf{x}_{t-1},\boldsymbol{\Theta}) \propto \exp\left(-\lambda \textstyle\sum_{j=1}^p \kappa(\delta^2_{t,j})\right),$$

where

$$\kappa(z) := -2\ln\left(\int_0^\infty u^{n/2}e^{-uz/2}p_U(u)du\right), \forall z \geq 0, \tag{2}$$

and $\kappa'(z)$ is a completely monotone function [3]. Random vectors of the form $\mathbf{w}_t = \frac{\boldsymbol{\omega}_t}{\sqrt{\mathbf{u}_t}}$ are known as Normal/Independent [3]. Note that a choice of $\kappa(z) = \sqrt{z^2+\epsilon^2}$ results in the ϵ-perturbed Laplace distributions used in our model [2]. Given T observations $(\mathbf{y}_t)_{t=1}^T \in \mathbb{R}^{n_t}$ and conditionally independent samples $(\mathbf{x}_t)_{t=1}^T \in \mathbb{R}^p$, we denote the objective function of the MAP estimator by $\mathfrak{L}((\mathbf{x}_t)_{t=1}^T, \boldsymbol{\Theta})$, that is $\log\mathfrak{L}((\mathbf{x}_t)_{t=1}^T, \boldsymbol{\Theta}) = \sum_{t=1}^T \log p(\mathbf{y}_t|\mathbf{x}_t,\boldsymbol{\Theta}) + \log p(\mathbf{x}_t|\mathbf{x}_{t-1},\boldsymbol{\Theta})$. Consider the current estimates $\left\{(\widehat{\mathbf{x}}_t^{(l)})_{t=1}^T, \widehat{\boldsymbol{\Theta}}^{(l)}\right\}$ at iteration l. Then:

$$\begin{aligned} &\log\mathfrak{L}((\mathbf{x}_t)_{t=1}^T, \boldsymbol{\Theta}) - \sum_{t=1}^T \log p(\mathbf{y}_t|\mathbf{x}_t,\boldsymbol{\Theta}) \\ &= \sum_{t,j=1}^{T,p} \log\left(\int_{(\mathbf{u}_t)_j} p\left((\boldsymbol{\omega}_t)_j, (\mathbf{u}_t)_j|\boldsymbol{\Theta}\right) d(\mathbf{u}_t)_j\right) \\ &= \sum_{t,j=1}^{T,p} \log\left(\int_{(\mathbf{u}_t)_j} \frac{p\left((\mathbf{u}_t)_j \middle| \left((\widehat{\mathbf{x}}_t^{(l)})_j\right)_{t=1}^T, \widehat{\boldsymbol{\Theta}}^{(l)}\right)}{p\left((\mathbf{u}_t)_j \middle| \left((\widehat{\mathbf{x}}_t^{(l)})_j\right)_{t=1}^T, \widehat{\boldsymbol{\Theta}}^{(l)}\right)} p\left((\boldsymbol{\omega}_t)_j, (\mathbf{u}_t)_j|\boldsymbol{\Theta}\right) d(\mathbf{u}_t)_j\right) \\ &\geq \sum_{t,j=1}^{T,p} \int_{(\mathbf{u}_t)_j} p\left((\mathbf{u}_t)_j \middle| \left((\widehat{\mathbf{x}}_t^{(l)})_j\right)_{t=1}^T, \widehat{\boldsymbol{\Theta}}^{(l)}\right) \log\left(\frac{p\left((\boldsymbol{\omega}_t)_j, (\mathbf{u}_t)_j|\boldsymbol{\Theta}\right)}{p\left((\mathbf{u}_t)_j \middle| \left((\widehat{\mathbf{x}}_t^{(l)})_j\right)_{t=1}^T, \widehat{\boldsymbol{\Theta}}^{(l)}\right)}\right) d(\mathbf{u}_t)_j \\ &= \sum_{t,j=1}^{T,p} \mathbb{E}_{(\mathbf{u}_t)_j \left| \left((\widehat{\mathbf{x}}_t^{(l)})_j\right)_{t=1}^T, \widehat{\boldsymbol{\Theta}}^{(l)}\right.} \left\{\log p((\boldsymbol{\omega}_t)_j, (\mathbf{u}_t)_j|\boldsymbol{\Theta})\right\} + C, \end{aligned} \tag{3}$$

where the inequality follows from Jensen's inequality and the constant C accounts for terms which do not depend on $\boldsymbol{\Theta}$. The so called Q-function is defined as:

$$\begin{aligned} &Q\left((\mathbf{x}_t)_{t=1}^T, \boldsymbol{\Theta}\middle| \left(\widehat{\mathbf{x}}_t^{(l)}\right)_{t=1}^T, \widehat{\boldsymbol{\Theta}}^{(l)}\right) := \sum_{t=1}^T \log p(\mathbf{y}_t|\mathbf{x}_t,\boldsymbol{\Theta}) \\ &+ \sum_{t,j=1}^{T,p} \mathbb{E}_{(\mathbf{u}_t)_j \left| \left((\widehat{\mathbf{x}}_t^{(l)})_j\right)_{t=1}^T, \widehat{\boldsymbol{\Theta}}^{(l)}\right.} \left\{\log p((\boldsymbol{\omega}_t)_j, (\mathbf{u}_t)_j|\boldsymbol{\Theta})\right\}. \end{aligned} \tag{4}$$

The EM algorithm maximizes the lower bound given by the Q-function of (4) instead of the log-likelihood itself. Moreover for all $t \in [T], j \in [p]$ and $\kappa(z) = \sqrt{z^2+\epsilon^2}$ we have [3]:

$$\mathbb{E}_{(\mathbf{u}_t)_j \left| \left((\widehat{\mathbf{x}}_t^{(l)})_j\right)_{t=1}^T, \widehat{\boldsymbol{\Theta}}^{(l)}\right.} \left\{\log p((\boldsymbol{\omega}_t)_j, (\mathbf{u}_t)_j|\boldsymbol{\Theta})\right\} = -\frac{\lambda}{2}\frac{(\mathbf{x}_t - \boldsymbol{\Theta}\mathbf{x}_{t-1})^2_j + \epsilon^2}{\sqrt{(\widehat{\mathbf{x}}_t^{(l)} - \widehat{\boldsymbol{\Theta}}^{(l)}\widehat{\mathbf{x}}_{t-1}^{(l)})^2_j + \epsilon^2}},$$

which after replacement results in the state-space model given by Eq. (9). This expectation gets updated in the outer EM loop using the *final* outputs of the inner loop. The outer EM algorithm can thus be summarized as forming the Q-function (E-step) and maximizing over $\boldsymbol{\Theta}$ (M-step), which is known to converge to a stationary point due to its ascent property [2]. As discussed in Section II-B, the outer M-step is implemented by another instance of the EM algorithm by alternating between Fixed Interval Smoothing (E-step) and updating $\boldsymbol{\Theta}$ (M-step).

REFERENCES

[1] X. Meng, J. P. Kao, H.-K. Lee, and P. O. Kanold, "Visual deprivation causes refinement of intracortical circuits in the auditory cortex," *Cell reports*, vol. 12, no. 6, pp. 955–964, 2015.

[2] D. Ba, B. Babadi, P. L. Purdon, and E. N. Brown, "Convergence and stability of iteratively re-weighted least squares algorithms," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 183–195, 2014.

[3] K. Lange and J. S. Sinsheimer, "Normal/independent distributions and their applications in robust regression," *Journal of Computational and Graphical Statistics*, vol. 2, no. 2, pp. 175–198, 1993.